

Transcriptional Analysis of Left-sided Colitis, Pancolitis, and Ulcerative Colitis-associated Dysplasia

Jacob T. Bjerrum, MD, PhD,^{*,†} Ole H. Nielsen, DMSc,[†] Lene B. Riis, MD, PhD,[‡] Valerie Pittet, PhD,[§] Christoph Mueller, PhD,^{||} Gerhard Rogler, DMSc,[¶] and Jørgen Olsen, DMSc^{*}

Background: It is unknown why patients with extensive ulcerative colitis (UC) have a higher risk of colorectal cancer compared with patients with left-sided UC. This study characterizes the inflammatory processes in left-sided UC, pancolitis, and UC-associated dysplasia at the transcriptional level to identify potential biomarkers and transcripts of importance for the carcinogenic behavior of chronic inflammation.

Methods: The Affymetrix GeneChip Human Genome U133 Plus 2.0 was applied on colonic biopsies from UC patients with left-sided UC, pancolitis, dysplasia, and controls. Reverse transcription polymerase chain reaction and immunohistochemistry were performed for validating selected transcripts in the initial cohort and in 2 independent cohorts of patients with UC. Microarray data were analyzed by principal component analysis, and reverse transcription polymerase chain reaction and immunohistochemistry data by the Wilcoxon's rank-sum test.

Results: The principal component analysis results revealed separate clusters for left-sided UC, pancolitis, dysplasia, and controls. Close clustering of dysplastic and pancolitic samples indicated similarities in gene expression. Indeed, 101 and 656 parallel upregulated and downregulated transcripts, respectively, were identified in specimens from dysplasia and pancolitis. Validation of selected transcripts hereof identified insulin receptor alpha (INSRA) and MAP kinase interacting serine/threonine kinase 2 (MKNK2) with an enhanced expression in dysplasia compared with left-sided UC and controls, whereas laminin γ 2 (LAMC2) was found with a lower expression in dysplasia compared with the remaining 3 groups.

Conclusions: This study demonstrates pancolitis and left-sided UC as distinct inflammatory processes at the transcriptional level, and identifies INSRA, MKNK2, and LAMC2 as potential critical transcripts in the inflammation-driven preneoplastic process of UC.

(*Inflamm Bowel Dis* 2014;20:2340–2352)

Key Words: gene expression, inflammatory bowel disease, insulin receptor alpha, laminin gamma 2, MAP kinase interacting serine/threonine kinase 2

Ulcerative colitis (UC)¹ and Crohn's disease (CD)² are the 2 most prevailing chronic inflammatory bowel diseases (IBD) emerging as global diseases with an increasing incidence and prevalence worldwide.³ UC and colonic CD are associated with

a higher risk of colorectal cancer (CRC),⁴ and the typical pathological changes progress from quiescent disease to active chronic inflammation without dysplasia to dysplasia, and ultimately cancer.⁵ It is assumed that this neoplastic process is inflammation driven,⁶ which is supported by recent studies^{7–9} identifying a continuous inflammatory state in the colonic mucosa of patients with UC, but not in CD, revealing a plausible explanation as to why the risk of CRC is higher in UC than CD. Hence, the chronic inflammatory load seems to be the central element for an increased risk of UC-associated CRC.^{10,11} Furthermore, a range of epidemiological studies have repeatedly recognized pancolitis to be an independent risk factor for CRC compared with less widespread colonic inflammation, e.g., left-sided UC.^{12,13} However, it is unknown why the inflammatory process in pancolitis leads to a higher risk of CRC compared with left-sided UC. One hypothesis is that the risk is increased because of an enhanced amount of epithelial cells in the pancolitic mucosa exposed to carcinogenic insults generated by the inflammation as compared with left-sided UC, a theory in line with the above-mentioned inflammatory load. Alternatively, the inflammation per se may be more "carcinogenic" in the pancolitis mucosa, and if so, an opportunity exists for the development of more rational treatment principles for pancolitis. Thus, the question is whether the inflammation of left-sided UC and pancolitis is identical or of different and gradually more carcinogenic types.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ibdjournal.org).

Received for publication August 4, 2014; Accepted August 29, 2014.

From the *Department of Cellular and Molecular Medicine, the Panum Institute, University of Copenhagen, Copenhagen, Denmark; [†]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark; [‡]Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark; [§]Health Care Evaluation Unit, Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland; ^{||}Division of Experimental Pathology, Institute of Pathology, Division of Immunopathology, University of Bern, Bern, Switzerland; and [¶]Division of Gastroenterology and Hepatology, University Hospital Zürich, Zürich, Switzerland.

Supported by grants from the Danish Agency for Science, Technology and Innovation (Grant No. 10-094201/FSS), the Lundbeck Foundation (Grant No. R17-A1597), and the Swiss IBD Cohort Study (SIBDC), funded by the Swiss National Science Foundation (SNSF) (Grant No. 3347CO-108792).

The authors have no conflicts of interest to disclose.

Reprints: Jacob T. Bjerrum, MD, PhD, Department of Gastroenterology, Medical Section, Herlev Hospital, Herlev Ringvej 75, 2730 Herlev, Denmark (e-mail: bjerrum.jacob@gmail.com).

Copyright © 2014 Crohn's & Colitis Foundation of America, Inc.

DOI 10.1097/MIB.0000000000000235

Published online 28 October 2014.

Therefore, the aim of this study was to use microarray-based gene expression profiling of colonic mucosal biopsies from controls and patients with UC to characterize the inflammatory processes in left-sided UC, pancolitis, and UC-associated dysplasia and subsequently to identify potential biomarkers and transcripts of importance for the carcinogenic behavior of chronic inflammation.

MATERIALS AND METHODS

Patient Population

Cohort 1, the Microarray Cohort

Colonic mucosal biopsies were from patients from the Department of Gastroenterology, Medical Section, Herlev Hospital, Denmark or from the Swiss IBD Cohort Study (SIBDCS), Switzerland. Sixty patients were enrolled: 6 patients with UC with dysplasia, 19 patients with UC with pancolitis, and 20 UC patients with left-sided colitis, as well as 15 controls, i.e., patients undergoing colonoscopy and where all investigations subsequently turned out to be normal (Table 1).

Cohort 2, the Polymerase Chain Reaction Cohort

Colonic mucosal biopsies were from patients, subjected to flexible colonoscopy at the Department of Gastroenterology, Medical Section, Herlev Hospital, Denmark. Forty patients were enrolled: 23 UC patients with left-sided colitis and 17 patients with UC with pancolitis (Table 1).

Cohort 3, the Polymerase Chain Reaction and Immunohistochemistry Cohort

The third cohort was established through inclusion of formalin fixed paraffin embedded (FFPE) colonic biopsies originating from the Department of Pathology, Herlev Hospital, Denmark. Six patients with UC with dysplasia, 4 patients with UC with pancolitis, and 4 UC patients with left-sided UC were included (Table 1).

All eligible patients with UC had their diagnosis verified by established criteria,¹⁴ and when biopsies were obtained, they were graded in accordance with the Mayo score.¹⁵ The endoscopic assessment of activity was confirmed by a histological examination, and any discrepancy between the observations led to exclusion of the specific patient. The included controls had no signs of inflammation. Exclusion criteria were age above 80 or below 18 years, clinical evidence of infection, recent (within 14 d) intake of antibiotics or probiotics, pregnancy, or severe mental illness.

Ethical Considerations

The Swiss Ethics Committees approved the SIBDCS, and the Scientific Ethics Committee of the Copenhagen Region (H-KA-20060164) approved the study as well. All patients participated in this study in confirmation with the principals outlined in the Declaration of Helsinki. Informed written consent was obtained from each participating person.

Tissue Samples

Cohorts 1 and 2

Mucosal colonic biopsies (10–20 mg each) were obtained endoscopically using routine endoscopic forceps. The descending colon was explicitly chosen to avoid any intersegmental variation in gene expression and because this is the most frequent area of inflammation. The biopsies originated from an inflamed area (aside from the control group and 3 of the 6 dysplastic samples) without ulceration. The first biopsy was used for a histopathological evaluation conducted in an unblinded fashion by staff pathologists in accordance with well-established criteria and focused on confirming the presence or absence of inflammation.¹⁶ The second biopsy was immediately stabilized in RNAlater (Ambion, Austin, TX) to minimize RNA degradation, and after 48 hours of storage in RNAlater at 4°C, the biopsy was kept at –80°C.

All dysplastic samples (cohort 1) also originated from the left side of the colon and contained low-grade dysplasia except for one sample containing high-grade dysplasia.

Cohort 3

FFPE biopsies from the descending colon of UC patients with left-sided UC, pancolitis, and dysplasia were randomly chosen from the biobank at the Department of Pathology, Herlev Hospital, Denmark. Just as in cohorts 1 and 2, the biopsies originated from an inflamed area without ulceration, which was confirmed in a blinded fashion by staff pathologists. Four of the dysplastic samples contained low-grade dysplasia, whereas 2 contained high-grade dysplasia (Table 1).

RNA Extraction, Hybridization, Detection, and Quantification of Signals

RNA was extracted from biopsies in RNAlater with the NucleoSpin RNA/Protein mini kit (Macherey-Nagel, Düren, Germany) and from FFPE biopsies with Rneasy FFPE Kit (Qiagen, Hilden, Germany) in accordance with the manufactures protocols. Integrity and purity were verified with an Agilent Bioanalyzer (Palo Alto, CA).

In accordance with the Affymetrix protocol and the one-cycle eukaryotic target labeling assay, biotin-labeled complementary RNA was produced. The complementary RNA was fragmented, and a hybridization mix was prepared, which included the fragmented target, probe array controls, bovine serum albumin, and herring sperm DNA. In this experiment, the Affymetrix GeneChip Human Genome U133 Plus 2.0 was applied (>47,000 transcripts). The hybridized probe array was subsequently stained with fluorescent protein streptavidin–phycoerythrin¹⁷ and scanned with a GeneArray scanner at the excitation wavelength of 488 nm. The amount of light emitted at 570 nm was proportional to the bound target at each location on the probe array. Data were stored as image files for further analysis. These procedures were all performed at the Microarray Core Facility, Rigshospitalet, Copenhagen, Denmark. The samples were run as 2 batches containing a mixture of sample types, i.e., pancolitis, left-sided UC, dysplasia, and controls, facilitating

TABLE 1. Clinical Details

	Dysplasia	Pancolitis	Left-sided UC	Controls	<i>P</i>
Cohort 1 (microarray)					
n	6	19	20	15	
Gender (male/female)	4/2	9/10	7/13	4/11	0.32
Age, median (IQR), yr	49 (28)	32 (22)	42 (30)	52 (34)	0.16
Years with disease (<10 yr/>10 yr)	2/4	13/6	13/7	—	0.29
Mayo score, median (IQR)	1.5 (5)	6 (5)	6 (4)	—	0.07
Mayo endoscopic score, median (IQR)	0.5 (2)	2 (1)	2 (0)	—	0.09
Smoking/nonsmoking	0/6	0/19	3/17	2/13	0.27
Daily medication ^a	2/0/0/0/2/0/2	15/5/6/2/8/5/0	17/3/3/2/5/1/2	—	0.09
Cohort 2 (RT-PCR)					
n	—	17	23	—	
Gender (male/female)	—	6/11	10/13	—	0.82
Age, median (IQR), yr	—	32 (17)	40 (17)	—	0.03
Years with disease (<10 yr/>10 yr)	—	12/5	16/7	—	1.0
Mayo score, median (IQR)	—	11 (4)	6 (2)	—	0.0001
Mayo endoscopic score, median (IQR)	—	2 (1)	1 (1)	—	0.0001
Smoking/nonsmoking	—	3/14	0/23	—	0.09
Daily medication ^a	—	15/8/5/1/5/2/2	18/15/5/1/5/1/2	—	0.93
Cohort 3 (RT-PCR and IHC)					
n	6	4	4	—	
Gender (male/female)	3/3	3/1	3/1	—	0.62
Age, median (IQR), yr	61 (37)	51 (18)	47 (35)	—	0.57
Years with disease (<10 yr/>10 yr)	1/5	4/0	3/1	—	0.02
Mayo score, median (IQR)	4 (5)	6.5 (2)	6 (5)	—	0.42
Mayo endoscopic score, median (IQR)	1 (2)	2 (1)	1.5 (1)	—	0.30
Smoking/nonsmoking	0/6	0/4	0/4	—	na
Daily medication ^a	4/1/0/0/2/0/1	2/1/0/0/0/0/2	2/1/1/0/1/0/1	—	0.77

Biopsies from cohorts 1 and 2 are fresh Later-conserved biopsies, whereas biopsies from cohort 3 are FFPE biopsies.

^aSystemic 5-ASA/topical 5-ASA/systemic steroids/topical steroids/thiopurines/infliximab/none.

IQR, interquartile range.

a subsequent identification of a potential batch effect with the principal component analysis (PCA). All data from this study have been submitted to the Gene Expression Omnibus (NCBI/ GEO: Series GSE47908), <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=rdrfuummasqyrc&acc=GSE47908>.

Real-time Reverse Transcription Polymerase Chain Reaction

Primers for amplification were chosen from the PrimerBank resource. All amplified polymerase chain reaction (PCR) products were sequenced to verify the expected identity, and melting curves were routinely checked to rule out the amplification of unrelated fragments during quantitative real-time reverse transcription polymerase chain reaction (RT-PCR). A 100 ng of reverse-transcribed total RNA was used as template. The PCR reactions were run using the LightCycler PCR system (Roche, Mannheim, Germany) as previously described.¹⁸ To standardize the

reactions, a serial 10-fold dilution of gel-purified PCR product of 1 mL was used as template in independent reactions, and the number of target molecules was calculated based on the standard curve. The number of ribosomal protein L10 transcripts (RPL10) was used for normalization between samples, as RPL10 was found constitutively and equally expressed in the microarray data from the 4 groups (calculated for probe set ID: “221989_at.” analysis of variance, $P = 0.99$) and thus a suitable positive PCR control. The amplification efficacy was calculated from the slope of the standard curve for each primer pair. The average amplification efficiency for the primer pairs was 86% (range, 75%–99%) for cohort 1, 88% (range, 76%–99%) for cohort 2, and 82% (range, 76%–99%) for cohort 3.

Immunohistochemistry

Four-micrometer sections of FFPE tissue were pretreated by deparaffinization, rehydration, and epitope retrieval in 1 step

by use of “Pretreatment Module” (PT-Link; Dako, Glostrup, Denmark) according to manufacturer’s instruction. Immunohistochemistry (IHC) staining was performed using the “Dako Autostainer Link” and with the commercially available kit (EnVision Flex; Dako). Slides were counterstained with Mayer’s hematoxylin before viewing. The antibodies were MAP kinase-interacting serine/threonine-protein kinase 2 (MKNK2, ab84345, dilution 1:500; Abcam), thyroid hormone receptor alpha 1 (THRA, ab53729, dilution 1:50; Abcam, Cambridge, MA), insulin receptor alpha (INSRA, I7661–25E, dilution 1:50; US Biological, Salem, MA), laminin gamma 2 (LAMC2, M 7262, dilution 1:50; Dako), integrin alpha 6 (ITGA6, HPA012696-100UL, dilution 1:50; Sigma–Aldrich, Brøndby, Denmark), and ezrin (EZR, sc-58758, dilution 1:50; Santa Cruz Biotechnology, Heidelberg, Germany). Spleen tissue was used as a positive control for MKNK2, and colon carcinoma tissue was used as positive control for THRA, LAMC2, ITGA6, and EZR. Placental tissue was used as a positive control for INSRA. Evaluation of the stained slides was performed by quantitative scoring of the intensity and fraction of cells stained. The immunoreactivity was graded from 0 to 3 according to intensity: 0 if the staining was absent, 1 for weak staining, 2 for moderate staining, and 3 for intense staining. The intensity scoring and evaluation of fraction of stained cells was performed blinded by an IBD pathologist.

Data Analysis

The clinical data (Table 1) were compared by the χ^2 test (categorical variables) or Kruskal–Wallis (continuous variables) with GraphPad Prism v.6.0 (GraphPad Software, La Jolla, CA). *P*-values <0.05 were considered significant.

Potential covariates in the clinical data (i.e., age, Mayo score, current medication, and years with disease) were accounted for with the use of partial least squares regression-discriminant analysis in SIMCA-P + 12 (Umetrics, Umeå, Sweden); microarray data as an X matrix and the potential covariates as the Y variables.¹⁹

By performing the robust multiarray analysis procedure²⁰ with quantile normalization implemented in the Affymetrix library for the R statistical environment,²¹ a single log₂ scale expression measure for each probe set was attained from the low-level data files (CEL files).

The unsupervised multivariate data analysis tool PCA²² was used to determine whether any intrinsic clustering existed within the data set and to identify any obvious batch effects. This was performed with the R package *pcaGoPromoter*,²³ which provides an overview of the data by PCA and at the same time produces a functional interpretation by gene ontology (GO) terms (biological processes). The GO terms were included only if *P* < 0.005. In short, if different groups of samples are clearly separated along a principal component axis, the *pcaGoPromoter* algorithm will identify the transcripts holding the strongest differential power and subject the identified transcripts to an overrepresentation analysis, e.g., a Fisher’s exact test.

The Hotelling’s *T*² test was conducted on the 2-component PCA model. In brief, the scores for each sample were applied to a multiple linear regression model using a dummy variable to

encode the classes (left-sided UC, pancolitis, and dysplasia, respectively). The calculations were performed in R. Dysplasia was tested against pancolitis and left-sided UC, respectively.

To determine whether the inflammation per se may be more carcinogenic in the pancolitis mucosa compared with left-sided UC, corresponding deregulated transcripts had to be identified in both pancolitis and dysplasia when compared with left-sided UC and controls. Thus, to extract a list of genes with higher expression values in dysplasia and pancolitis as compared with left-sided UC, several filters were applied: (1) The false discovery rate corrected *P*-value of a *t* test (the expression data are on a log₂ scale, and a parametric test was accordingly used) comparing the mean expression in dysplasia with left-sided UC should be <0.0001, and the fold differences 2-fold or greater in dysplasia, arbitrary values. (2) The mean difference in expression values between dysplasia and pancolitis should be less than 1.3 (on a normal scale). This filter assures that the upregulated transcripts found in dysplasia versus left-sided UC are also upregulated in pancolitis. (3) The mean difference in expression values between left-sided UC and controls should be less than 1.3 (on a normal scale). This filter eliminates transcripts that just reflect inflammation per se.

To extract a list of transcripts with lower expression in dysplasia compared with left-sided UC the following filter was applied: The false discovery rate corrected *P*-value of a *t* test comparing the mean expression in dysplasia with the mean expression in left-sided UC should be <0.0001, and the mean expression in left-sided UC should be more than 2.5-fold higher (on a normal scale) than the mean expression level in dysplasia.

The RT-PCR data and IHC data were analyzed with the Wilcoxon’s rank-sum test, and the results were considered significant when *P* < 0.05.

RESULTS

Array Analysis Reveals Separate Clusters of Gene Expression for Left-sided UC and Colitis, Pancolitis, Dysplasia, and Controls

An initial PCA score plot containing all 4 sample types, i.e., left-sided UC, pancolitis, dysplasia, and controls, was performed to make sure that the diseased samples clustered away from the control samples and to detect any potential batch effect. As seen in Figure 1A, there is a separate clustering of all 4 groups, which means that the gene expression profiles of the diseased samples are distinct from each other and the control samples. The clustering of the 4 groups also implies that the batch effect is at an absolute minimum as the samples were run in 2 batches with a mixture of sample types. To get insight into the biological processes involved and to identify differentially expressed transcripts, a subsequent PCA model was constructed containing only diseased samples. This PCA score plot (Fig. 1B) illustrates a 3-way separation of the samples in accordance with class belongings: left-sided UC, pancolitis, and dysplasia, respectively.

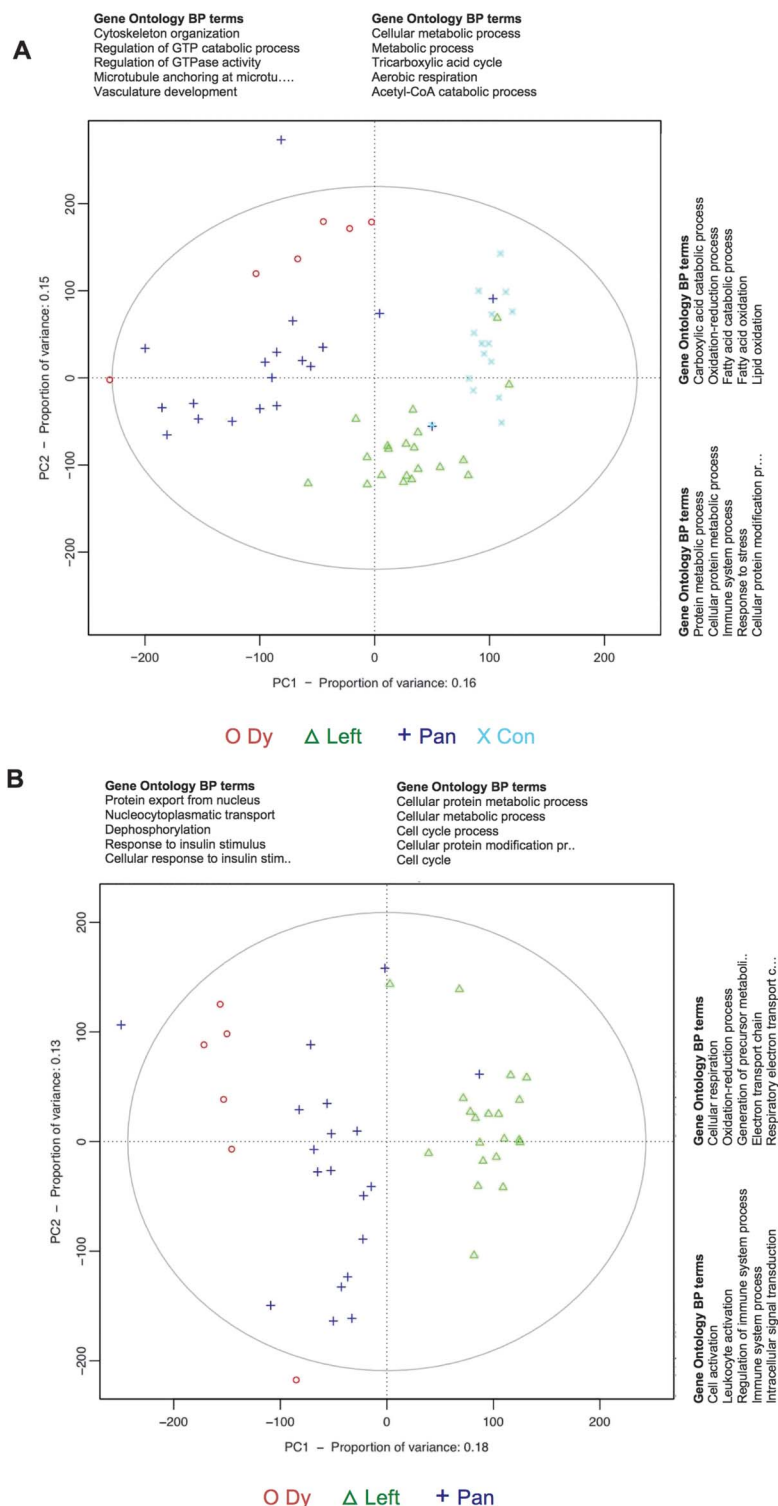


FIGURE 1. PCA score plots of the microarray data. A, The PCA score plot is described by 2 principal components (PC1 and PC2) containing 31% of the variance and illustrates separate clustering of the 4 sample types with only very few samples overlapping. These samples have been re-evaluated without any changes in their diagnosis. B, The PCA score plot illustrates a 3-way separation of the samples in accordance with class belonging. The separation is primarily along PC1 containing 18% of the variance: samples from patients with left-sided UC are localized toward positive values, samples from patients with dysplasia toward negative values, and samples from patients with pancolitis in the middle. Each axis in both PCA score plots is functionally annotated with the 5 most significant ($P < 0.005$) Gene ontology terms. BP, biological processes.

The PCA model is described by 2 components explaining a total of 31% (R^2X) of the variation in the dataset. Most of the variance is observed along the first principal component (PC1, $R^2X = 18\%$) with samples from patients with left-sided UC located toward positive values, dysplasia toward negative values, and pancolitis in the middle. The Hotelling's T^2 test was conducted on the PCA model and confirmed significant differences of the gene expression signatures: dysplasia versus pancolitis, $P < 0.01$; dysplasia versus left-sided UC, $P < 10^{-13}$; and pancolitis versus left-sided UC, $P < 10^{-9}$.

To test whether potential covariates could have a significant effect on these results, individual partial least squares regression-discriminant analysis models were constructed for age, years with disease, Mayo score, Mayo endoscopic score, and medication, as these were either significantly different between the groups or a trend was observed (Table 1). However, none of these models turned out to be predictive (data not shown).

Identification of Transcripts Relevant for Clustering

The top 5 GO terms that relate to the gene expression profiles of left-sided UC are described by processes such as *protein metabolic process* and *cell cycle process*, whereas the GO terms that represent dysplasia and pancolitis relate to biological processes such as *cellular response to insulin* and *response to insulin stimulus* (Fig. 1B).

To identify the gene expression profile responsible for the differentiation between pancolitis and left-sided UC and to simultaneously identify potential biomarkers or transcripts of significant importance in the inflammation-driven preneoplastic process, the statistical filters described in Materials and Methods were applied. The resulting lists of deregulated transcripts are given in Tables, Supplemental Digital Content 1 and 2, <http://links.lww.com/IBD/A625> and <http://links.lww.com/IBD/A626>, respectively. Table, Supplemental Digital Content 1, <http://links.lww.com/IBD/A625> contains 101 transcripts significantly upregulated in both pancolitis and dysplasia compared with left-sided UC and controls, whereas Table, Supplemental Digital Content 2, <http://links.lww.com/IBD/A626> represents 656 significantly downregulated transcripts in pancolitis and dysplasia compared with left-sided UC and controls.

Validation of Relevant Transcripts by RT-PCR and IHC

A panel of selected transcripts from these lists was internally validated by RT-PCR on all samples from cohort 1 (the microarray cohort, i.e., 6 dysplasia, 19 pancolitis, 20 left-sided UC, and 15 controls; Table 1). As it is extremely difficult to get hold of fresh-frozen or RNAlater-conserved biopsies with UC-associated dysplastic material, the external validation of the selected transcripts was performed on 2 different external cohorts: cohort 2 (the PCR cohort, i.e., fresh samples from 23 left-sided UC and 17 pancolitis; Table 1) to focus on potential differences between pancolitis and left-sided UC and cohort 3 (the PCR and IHC cohort, i.e., FFPE biopsies from 4 pancolitis, 4 left-sided UC,

and 6 dysplasia; Table 1) to focus on potential differences between dysplasia, pancolitis, and left-sided UC and evaluate the potential significance of FFPE biopsies as validation material.

The selection of transcripts from these lists was based on their known involvement in inflammatory and neoplastic processes, and they had to be related to 1 or more of the biological processes described in the PCA score plots. RT-PCR was consequently performed in cohort 1 on insulin receptor alpha (INSRA), thyroid hormone receptor alpha 1 (THRA), MAP kinase-interacting serine/threonine-protein kinase 2 (MKNK2), laminin gamma 2 (LAMC2), integrin alpha 6 (ITGA6), ezrin (EZR), epidermal growth factor receptor (EGFR), signal transducer and activator of transcription 1 (STAT1), integrin $\beta 6$ (ITGB6), cathepsin Z (CTSZ), hepatocyte nuclear factor 4 γ (HNF4G), apoptosis inhibitor 5 (API5), eukaryotic translation initiation factor 4A2 (EIF4A2), Sp3 transcription factor (SP3), myosin 1C (MYO1C), and laminin beta 1 (LAMB1). Transcripts without different expression levels in dysplasia and controls or within the diseased patient groups, i.e., dysplasia, pancolitis, and left-sided UC were excluded for further analysis (see Fig., Supplemental Digital Content 3, <http://links.lww.com/IBD/A627>), as the objective of this study was to identify potential biomarkers and transcripts of importance for the carcinogenic behavior of chronic inflammation. The RT-PCR results for which differences were observed are illustrated in Figures 2A and 3A. An overall higher expression of INSRA, THRA, and MKNK2 was seen in the diseased sample types (i.e., left-sided UC, pancolitis, and dysplasia) compared with controls (except for INSRA in left-sided UC). Furthermore, mimicking the microarray data, INSRA, THRA, and MKNK2 transcripts were found in higher abundance in dysplasia compared with left-sided UC (Fig. 2A). Similarly, a lower expression of LAMC2, ITGA6, and EZR was found in dysplasia compared with controls and inflamed tissue (except for EZR in pancolitis; Fig. 3A). These 6 transcripts were further sought validated in external validation cohorts.

Thus, focusing on the differences between pancolitis and left-sided UC cohort 2 demonstrated a significant difference in the expression of INSRA, whereas only a trend was observed for MKNK2, and no difference was seen for THRA (Fig. 2B). Similarly, ITGA6 was found with a higher expression in pancolitis compared with left-sided UC, and trends were seen for LAMC2 and EZR (Fig. 3B).

In cohort 3, the expression of INSRA was significantly higher in dysplasia compared with left-sided UC, which was also true for MKNK2, where a significant difference was also seen between pancolitis and left-sided UC (Fig. 2C). Again, no differences were observed for THRA, although a clear trend was seen with higher expression in dysplasia compared with left-sided UC. Finally, no differences between the 3 groups could be identified in relation to LAMC2, ITGA6, and EZR, although P -values of 0.05 were seen with LAMC2 and ITGA6 when comparing pancolitis and left-sided UC (Fig. 3C).

To investigate whether the deregulated transcripts (i.e., MKNK2, THRA, INSRA, LAMC2, ITGA6, and EZR) translate

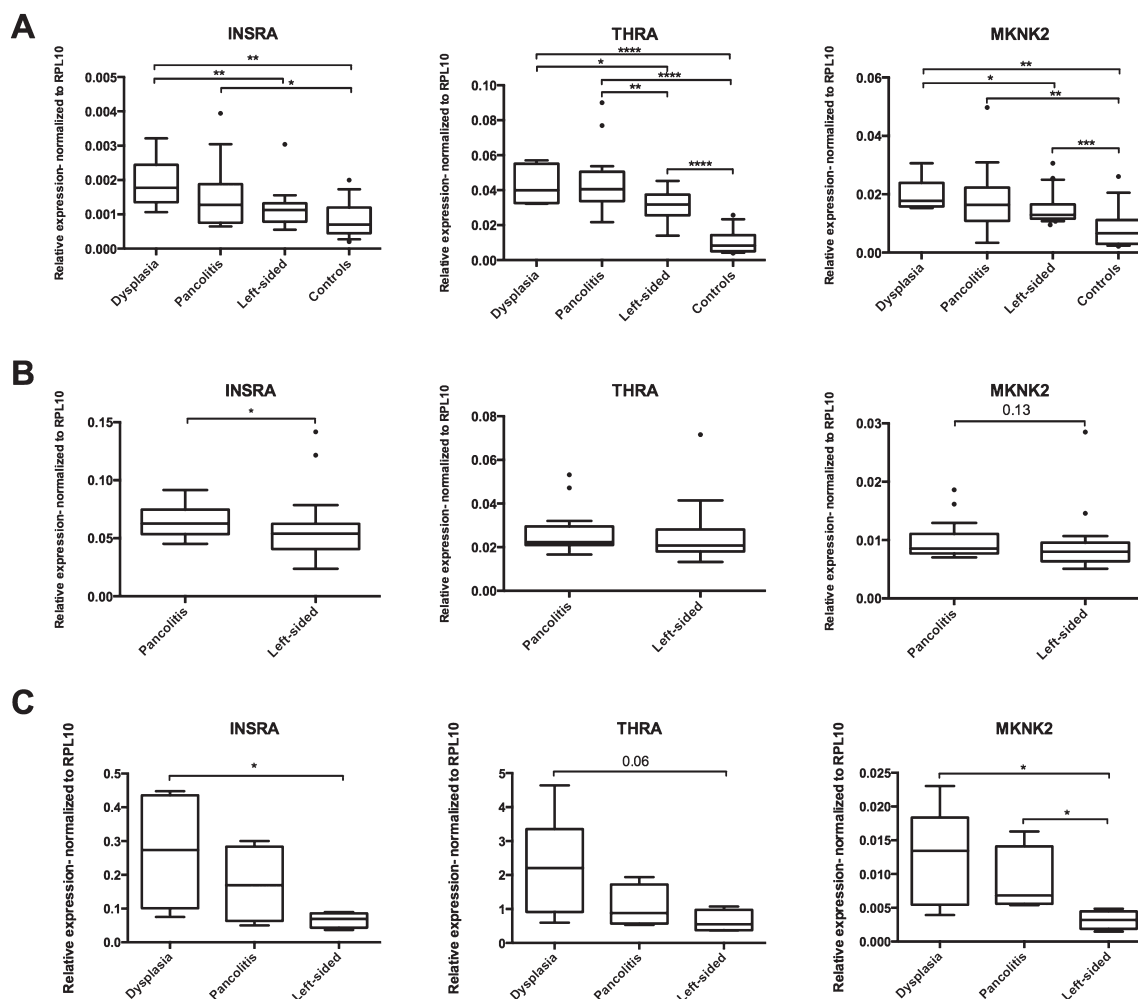


FIGURE 2. Real-time RT-PCR quantification of selected upregulated gene expressions. Transcripts for the insulin receptor alpha (INSRA), thyroid hormone receptor alpha 1 (THRA), and MAP kinase-interacting serine/threonine-protein kinase 2 (MKNK2) were measured in RNA extracted from (A) all of the samples in the initial microarray cohort 1, (B) the external validation cohort 2, and (C) external validation cohort 3. The number of ribosomal protein L10 (RPL10) transcripts was used for normalization between samples. The box border represents the interquartile range, and the horizontal line in the box is the median. The whiskers show the largest/smallest observation falling within a distance of 1.5 times the box size. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. The P -values are not corrected for false discovery rates.

into changed protein levels, IHC staining was performed on FFPE biopsies from all patients in cohort 3. As seen in Figures 4 and 5, a significantly increased intensity score was observed for INSR in dysplasia when compared with both pancolitis and left-sided UC, whereas the score was significantly lower for LAMC2 in dysplasia compared with both pancolitis and left-sided UC. No differences in the intensity scores could be documented for the remaining transcripts. Control hematoxylin and eosin of normal colonic mucosa, left-sided UC, pancolitis, and colonic mucosa with UC-associated dysplasia are presented in Fig., Supplemental Digital Content 4, <http://links.lww.com/IBD/A628>.

DISCUSSION

With the intention to characterize the inflammatory processes at the transcriptional level in left-sided UC, pancolitis, and

UC-associated dysplasia, this study used gene expression profiling of colonic mucosal biopsies. To the best of our knowledge, a similar approach has never previously been published, which makes the identification of pancolitis and left-sided UC as distinct inflammatory processes at the transcriptional level a novel finding (Fig. 1).

The transcriptional analysis also revealed that the gene expression profile of UC-associated dysplasia was distinct from both pancolitis and left-sided UC. However, the proximity of dysplastic and pancolitic samples in the PCA score plot (Fig. 1B) indicated similarities in the gene expression profile, and as the objective of this study was to identify transcripts of importance for the carcinogenic behavior of pancolitis, similarities between dysplasia and pancolitis were the main focus of the succeeding analyses. Thus, 101 and 656 parallel upregulated and downregulated transcripts, respectively, were subsequently identified in

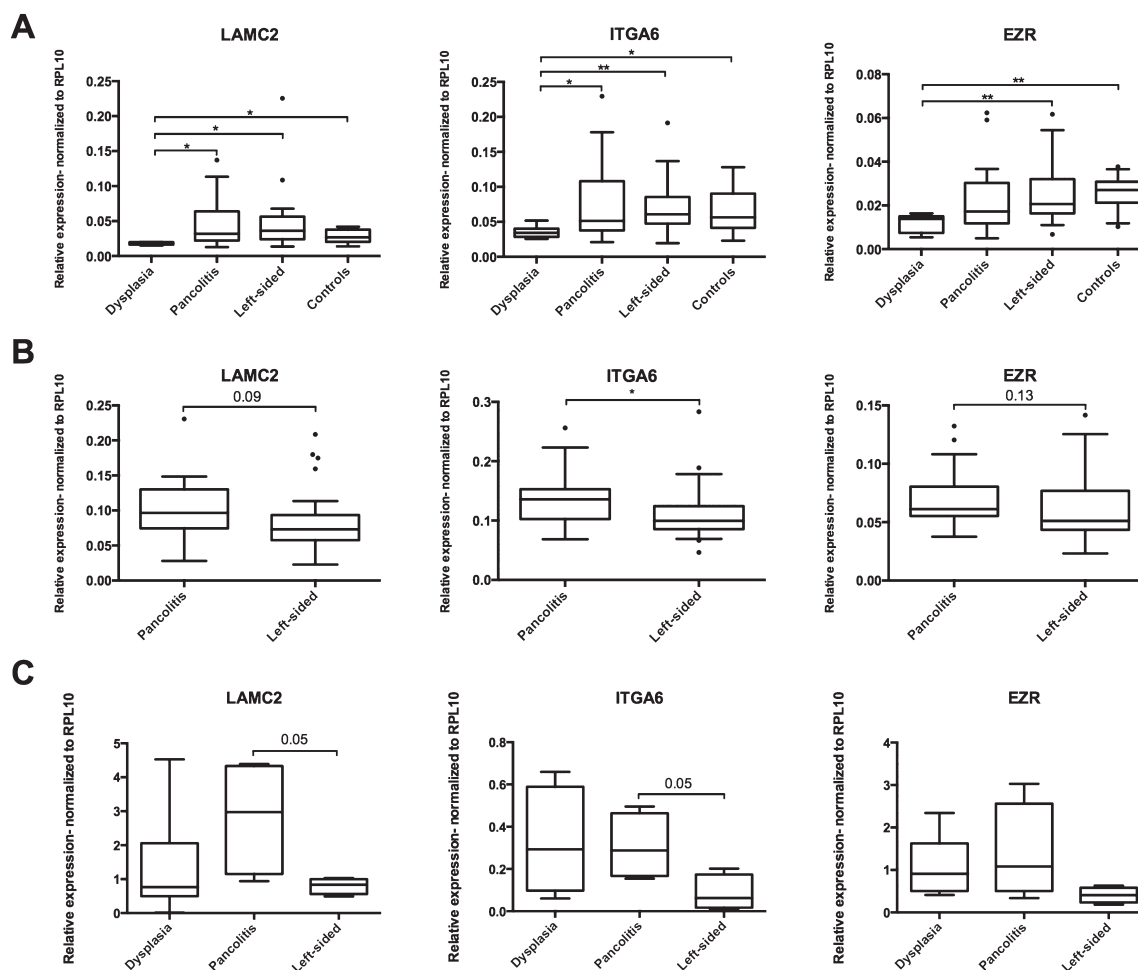


FIGURE 3. Real-time RT-PCR quantification of selected downregulated gene expressions. Transcripts for laminin gamma 2 (LAMC2), integrin alpha 6 (ITGA6), and ezrin (EZR) were measured in RNA extracted from (A) all of the samples in the initial microarray cohort 1, (B) the external validation cohort 2, and (C) external validation cohort 3. The number of ribosomal protein L10 (RPL10) transcripts was used for normalization between samples. The box border represents the interquartile range, and the horizontal line in the box is the median. The whiskers show the largest/smallest observation falling within a distance of 1.5 times the box size. * $P < 0.05$, ** $P < 0.01$. The P -values are not corrected for false discovery rates.

dysplasia and pancolitis (see Tables, Supplemental Digital Content 1 and 2, <http://links.lww.com/IBD/A625> and <http://links.lww.com/IBD/A626>).

From the upregulated list (see Table, Supplemental Digital Content 1, <http://links.lww.com/IBD/A625>), the RT-PCR validated transcripts of MKNK2, THRA, and INSRA showed an overall higher expression in diseased samples as compared with controls, and a significant difference was found between dysplasia and left-sided UC, which was not the case for pancolitis (cohort 1; Fig. 2A). These results were partially validated in cohorts 2 and 3, especially regarding INSRA and MKNK2, which were found highly expressed in dysplastic tissue as compared with left-sided UC, and MKNK2 was also found with a higher expression in pancolitis as compared with left-sided UC (cohorts 2 and 3; Fig. 2B, C).

The 3 transcripts, i.e., INSRA, THRA, and MKNK2 are part of the same intracellular signaling pathways: the

phosphatidylinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) signaling pathways (Fig. 6), which are key cellular transduction pathways in normal functional cells and important regulators of cellular growth, metabolism, and survival. Recent evidence, however, recognizes INSRA and THRA as being overexpressed in several tumors including colonic adenomas²⁴ and CRC.²⁵ The PI3K and MAPK signaling pathways converge on MKNK2, as does several pathways of a range of inflammation-associated cytokine receptors (e.g., interleukin 1 and tumor necrosis factor alpha; Fig. 6). Activated MKNK2 and MKNK1 (MKNKs) phosphorylate eukaryotic initiation factor 4E (eIF4E), which is the rate-limiting step in the activation of the eIF4G complex that directs protein biosynthesis. Thus, by inhibiting MKNKs, studies have demonstrated a significant reduction in the release of proinflammatory cytokines, such as tumor necrosis factor alpha, interleukin 1 β , and interleukin 6 in activated T cells and keratinocytes.^{26,27} Furthermore, several lines of evidence

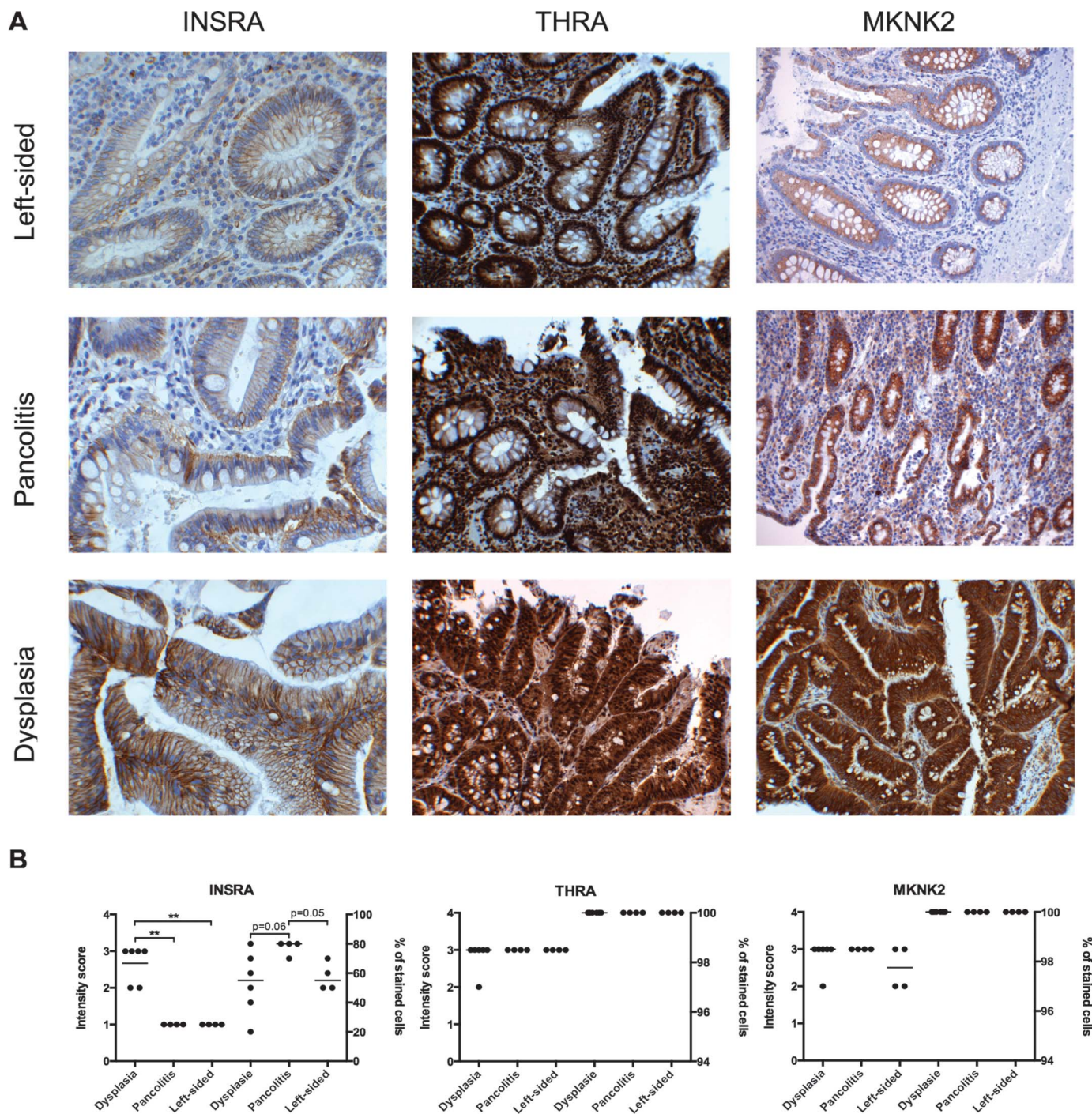


FIGURE 4. Immunohistochemical staining of INSRA, THRA, and MKNK2. A, Representative immunohistochemical staining of INSRA, THRA, and MKNK2 in FFPE colonic biopsies from a patient with dysplasia, pancolitis, or left-sided UC. The samples are counterstained with hematoxylin. Original magnification $\times 400$. INSRA is presented with a $\times 40$ lens to visualize the membrane staining, whereas THRA and MKNK2 are presented with a $\times 20$ lens. Thus, INSRA is expressed in the cytoplasm and along the basolateral membranes of the colonocytes, the THRA is expressed in the cytoplasm and nucleus, and MKNK2 is expressed primarily in the cytoplasm of colonocytes. B, Scatter plot of the intensity scores (0 = negative stain, 1 = mild stain, 2 = moderate stain, and 3 = strong stain) and fraction of stained cells for all 14 patients in cohort 3. The vertical lines represent the median scores. $**P < 0.01$. The P -values are not corrected for false discovery rates.

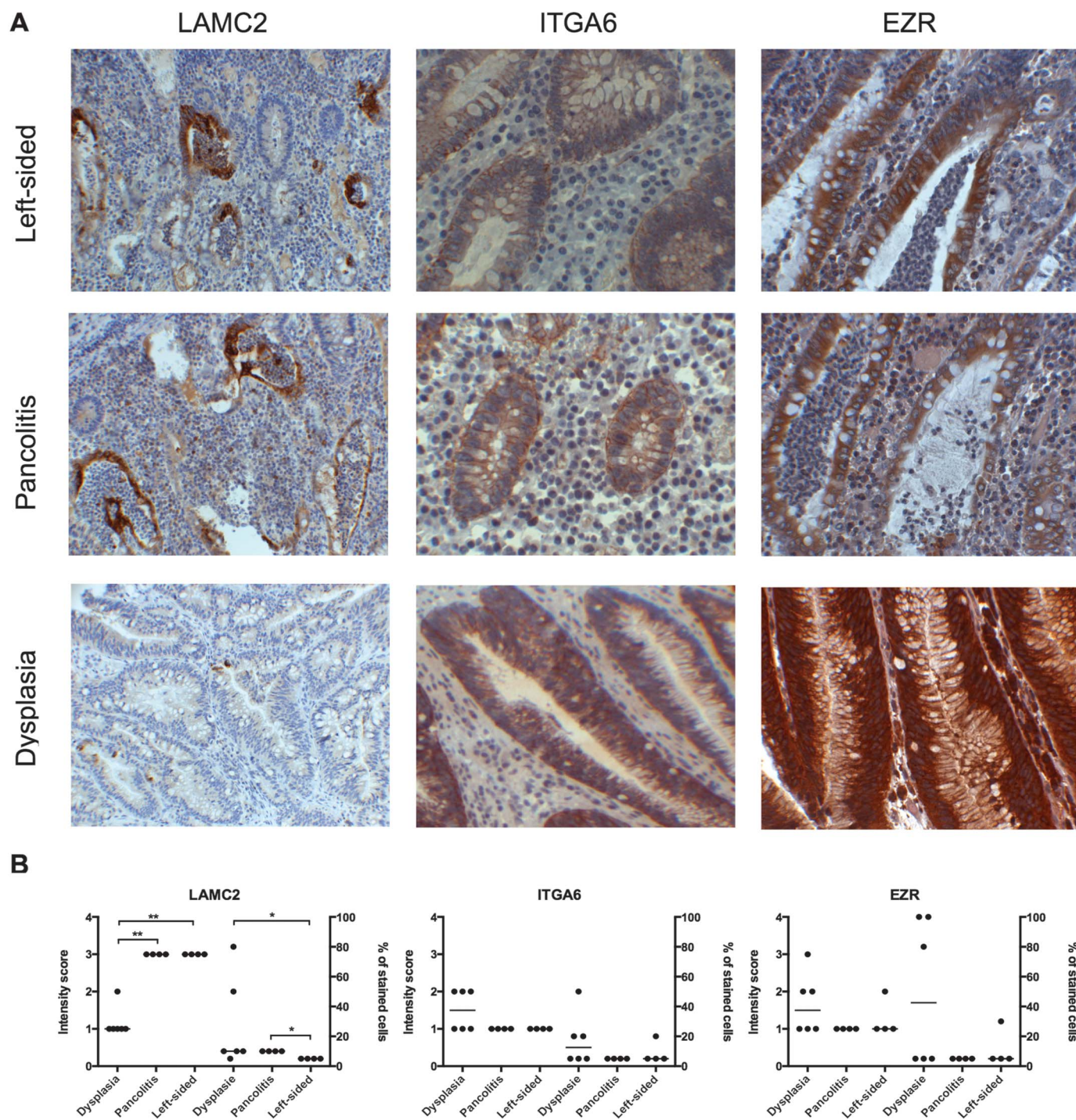


FIGURE 5. Immunohistochemical staining of LAMC2, ITGA6, and EZR. **A**, Representative immunohistochemical staining of LAMC2, ITGA6, and EZR in FFPE colonic biopsies from a patient with dysplasia, pancolitis, or left-sided UC. The samples are counterstained with hematoxylin. Original magnification $\times 400$. LAMC2 is presented with a $\times 20$ lens, whereas ITGA6 and EZR are presented with a $\times 40$ lens to visualize the membrane staining. Thus, LAMC2 is primarily expressed in the cytoplasm of colonocytes, but in regions with invasive appearances in the dysplastic tissue, the expression becomes more membranous. ITGA6 is expressed in the basolateral membrane of colonocytes and becomes more intense in regions with invasive characteristics in the dysplastic tissue. EZR is found primarily in the apical membrane of colonocytes but is also seen associated with the cytoskeleton. **B**, Scatter plot of the intensity scores (0 = negative stain, 1 = mild stain, 2 = moderate stain, and 3 = strong stain) and fraction of stained cells for all 14 patients in cohort 3. The vertical lines represent the median scores. * $P < 0.05$, ** $P < 0.01$. The P -values are not corrected for false discovery rates.

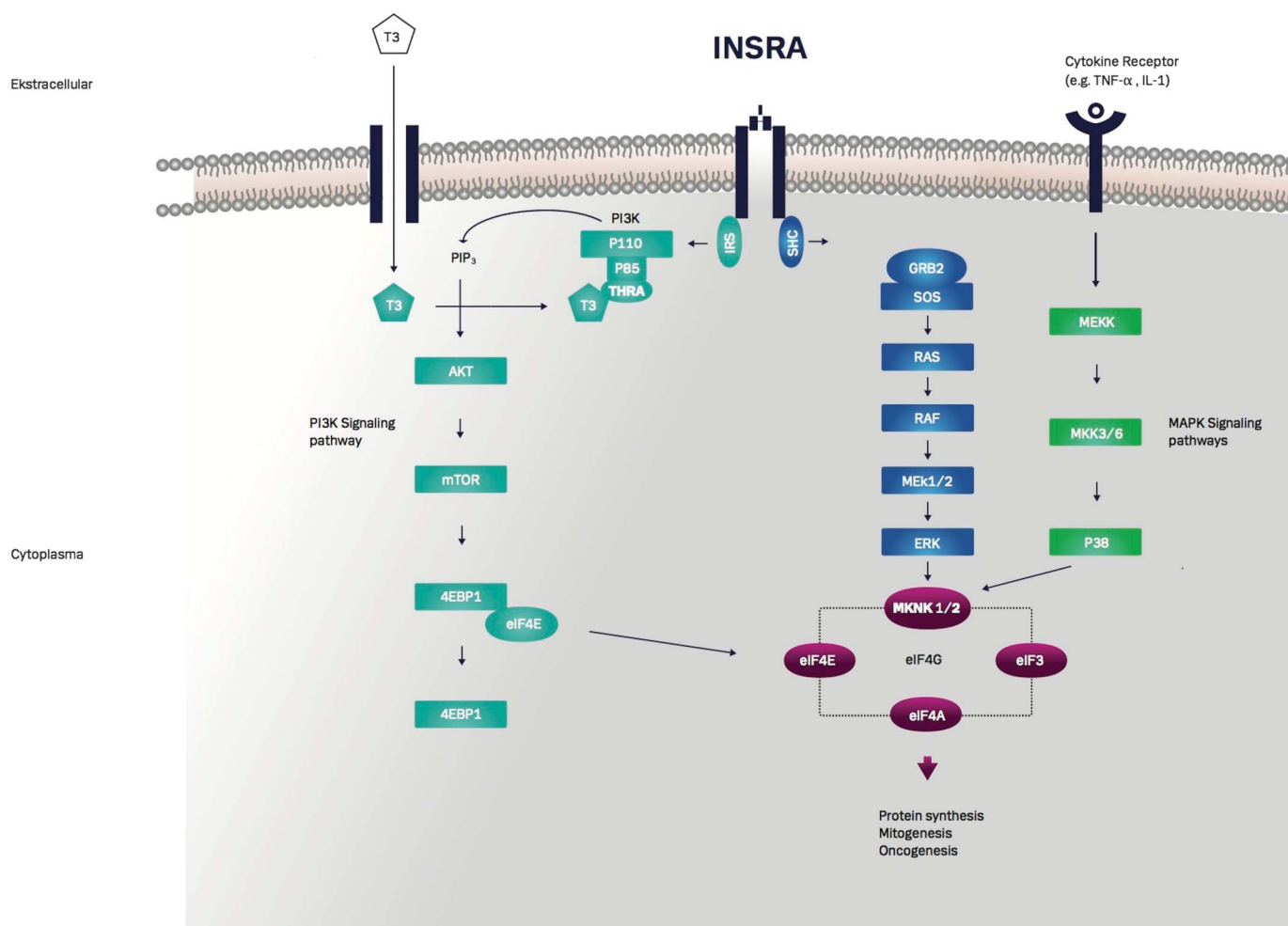


FIGURE 6. Schematic representation of the intracellular signaling pathways in which INSRA, THRA, and MKNK2 are pivotal elements. Ligand binding to INSRA stimulates the tyrosine kinase activity of the receptor, which phosphorylates several substrates including IRS1/2 (insulin receptor substrate 1/2) and Shc/Grb2/Sos (adaptor protein Src homology 2 domain-containing/shc/growth factor receptor bound protein-2/Son of Sevenless protein) complexes. IRS proteins and the ligand bound THRA are able to interact with the regulatory subunit (p85) of PI3K (phosphatidylinositol 3-kinase) resulting in an activation of the catalytic subunit (p110) and subsequently activation of AKT and mammalian target of Rapamycin (mTOR). Activated mTOR liberates eIF4E (eukaryotic initiation factor 4E) from the inhibitory binding protein 4EBP1 (eIF4E-binding protein 1), and eIF4E is free to engage the scaffolding protein eIF4G and be phosphorylated by MKNKs, which is the rate-limiting step in the activation of the eIF4F translation initiating complex (eIF4G, eIF4E, and eIF4A). The activated Shc/Grb2/Sos complex along with ligand bound cytokine receptors trigger the MAPK (mitogen-activated protein kinases) pathways of ERK (extracellular signal-regulated kinase) and p38, which activate MKNKs promoting protein synthesis, mitogenesis, and oncogenesis.

indicate that overexpressed and phosphorylated eIF4E serves as a genuine oncogenic accelerator in a milieu of cancer, which seems to play an important role in the early stages of malignant transformation as compared with the late stages of CRC.^{28–30} More importantly, a recent study confirmed that phosphorylation of eIF4E by MKNKs is essential for the ability of eIF4E to promote tumorigenesis,³¹ whereas it is dispensable for growth and development in normal tissue.³² Taken together, these observations make MKNKs ideal and novel targets in cancer therapy and in the treatment of IBD, and they may ultimately serve as a new therapeutic chemopreventive avenue in the treatment of especially pancolitis with potentially very few side effects.

From the downregulated list (see Table, Supplemental Digital Content 2, <http://links.lww.com/IBD/A626>), the RT-PCR validated transcripts of LAMC2, ITGA6, and EZR: showed a significant lower expression in dysplasia as compared with controls (cohort 1; Fig. 3) and inflamed samples (except for dysplasia versus pancolitis). No differences between inflamed samples were documented in cohort 1 (Fig. 3A), but external validation identified ITGA6 being significantly higher expressed in pancolitis compared with left-sided UC (cohort 2; Fig. 3B), and similar trends were observed for LAMC2 and EZR. Nevertheless, the proteins of all 3 transcripts are essential for the integrity and homeostasis of the epithelial barrier, as they are vital components

of various types of junctions (e.g., tight junctions, adherent junctions, and desmosomes) necessary for fundamental intercellular adhesion and cell–matrix interaction.^{33–35} In addition to their role in stable adhesion, these proteins are also important in cell migration and have consequently been found upregulated during wound healing and tumorigenesis.^{35–37}

The lower LAMC2, ITGA6, and EZR gene expressions in UC-associated dysplasia as described in cohort 1 have, to our knowledge, not been described previously. As they persistently have been found upregulated in CRC, the decreased expression seems to distinguish this premalignant state from both normal, inflamed, and cancerous tissue in the colon.^{38–41} Thus, especially LAMC2 may constitute a potential biomarker of preneoplastic lesions in patients with UC. However, the downregulation of LAMC2, ITGA6, and EZR seen in cohort 1 was not validated in cohort 3, which signifies one of the weaknesses in this study; cohort 3 was based on a relative low number of patients and FFPE biopsies, and the results have to be interpreted with this matter in mind. However, despite the expected low RNA integrity number values (average RNA integrity number value 2.1) acceptable amplification efficiencies (average 82%; range, 76%–99%) were achieved making the RT-PCR results valid and consequently illustrates the feasibility of using FFPE biopsies as a valuable source for retrospective gene expression studies in the future.

The findings at the transcriptional level concurred with the IHC results for INSRA and LAMC2. The INSRA demonstrated a significantly higher intensity score in dysplasia compared with inflamed samples (Fig. 4B), and LAMC2 a significantly lower intensity score in dysplasia compared with inflamed samples, whereas the fraction of stained cells was significantly higher in dysplasia and pancolitis as compared with left-sided UC (Fig. 5B). IHC did not indicate significant differences for expression of THRA, MKNK2, ITGA6, and EZR. If subtle differences are present at the protein level as indicated by the transcriptional data, more sensitive protein analyses are needed, which unfortunately was not an option in the current experimental setting. Further potential limitations of this study include the partial discrepancy between the microarray and PCR results, a discrepancy, i.e., explained by the 3' end based microarray. Correlation coefficients between these 2 technologies have been found to be as low as 0.4,⁴² but a possible solution to circumvent this problem in future studies would be the combination of microarray and cap analysis gene expression.⁴³ A higher number of dysplastic samples clearly would have been preferable, although exact power calculations are impossible without preliminary data at hand. Finally, a range of possible covariates was accounted for, but the potential confounding effect of these factors needs to be taken into consideration when interpreting the results.

In conclusion, with the use of gene expression profiles of mucosal biopsies, this study recognizes pancolitis and left-sided UC as distinct inflammatory processes at the transcriptional level and identifies transcripts of potential importance for the carcinogenic behavior of pancolitis. INSRA and MKNK2 seem to be key mediators of inflammation and tumorigenesis,

and especially MKNK2 presents itself as a possible new therapeutic chemopreventive avenue in the treatment of pancolitis. LAMC2 seems to constitute an ideal biomarker of preneoplastic lesions in patients with UC, although this remains to be validated. Finally, it has been demonstrated that FFPE biopsies are valid and a valuable source for retrospective gene expression studies.

Future studies at the transcription level will have to include inflammatory controls and non-IBD patients with CRC. At the functional level, studies will need to focus on 1 or a few of the identified transcripts linking pancolitis and UC-associated dysplasia, and thus evaluate whether manipulation with the expression levels will have any significant influence on experimental models of colitis and colitis-associated colon cancer carcinogenesis.

REFERENCES

- Ordás I, Eckmann L, Talamini M, et al. Ulcerative colitis. *Lancet*. 2012; 380:1606–1619.
- Baumgart DC, Sandborn WJ. Crohn's disease. *Lancet*. 2012;380:1590–1605.
- Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142:46–54.
- Jess T, Simonsen J, Jørgensen KT, et al. Decreasing risk of colorectal cancer in patients with inflammatory bowel disease over 30 years. *Gastroenterology*. 2012;143:375–381.
- Risques RA, Lai LA, Himmetoglu C, et al. Ulcerative colitis-associated colorectal cancer arises in a field of short telomeres, senescence, and inflammation. *Cancer Res*. 2011;71:1669–1679.
- Karin M, Greten FR. NF-kappaB: linking inflammation and immunity to cancer development and progression. *Nat Rev Immunol*. 2005;5:749–759.
- Planell N, Lozano JJ, Mora-Buch R, et al. Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. *Gut*. 2012;62:967–976.
- Olsen J, Gerds TA, Seidelin JB, et al. Diagnosis of ulcerative colitis before onset of inflammation by multivariate modeling of genome-wide gene expression data. *Inflamm Bowel Dis*. 2009;15:1032–1038.
- Bjerrum JT, Hansen M, Olsen J, et al. Genome-wide gene expression analysis of mucosal colonic biopsies and isolated colonocytes suggests a continuous inflammatory state in the lamina propria of patients with quiescent ulcerative colitis. *Inflamm Bowel Dis*. 2010;16:999–1007.
- O'Connor PM, Lapointe TK, Beck PL, et al. Mechanisms by which inflammation may increase intestinal cancer risk in inflammatory bowel disease. *Inflamm Bowel Dis*. 2010;16:1411–1420.
- Ullman TA, Itzkowitz SH. Intestinal inflammation and cancer. *Gastroenterology*. 2011;140:1807–1816.
- Lakatos L, Mester G, Erdelyi Z, et al. Risk factors for ulcerative colitis-associated colorectal cancer in a Hungarian cohort of patients with ulcerative colitis: results of a population-based study. *Inflamm Bowel Dis*. 2006;12:205–211.
- Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*. 2001;48:526–535.
- Nikolaus S, Schreiber S. Diagnostics of inflammatory bowel disease. *Gastroenterology*. 2007;133:1670–1689.
- Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med*. 1987;317:1625–1629.
- Bentley E, Jenkins D, Campbell F, et al. How could pathologists improve the initial diagnosis of colitis? Evidence from an international workshop. *J Clin Pathol*. 2002;55:955–960.
- Wang L, Lofton C, Popp M, et al. Using luminescent nanoparticles as staining probes for Affymetrix GeneChips. *Bioconjug Chem*. 2007;18: 610–613.
- Csillag C, Nielsen OH, Vainer B, et al. Expression of the genes dual oxidase 2, lipocalin 2 and regenerating islet-derived 1 alpha in Crohn's disease. *Scand J Gastroenterol*. 2007;42:454–463.

19. Eriksson L, Antti H, Gottfries J, et al. Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Anal Bioanal Chem*. 2004;380:419–429.
20. Irizarry RA, Bolstad BM, Collin F, et al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31:e15.
21. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5:R80.
22. Eriksson L, Andersson PL, Johansson E, et al. Megavariate analysis of environmental QSAR data. Part I—a basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). *Mol Divers*. 2006;10:169–186.
23. Hansen M, Gerds TA, Nielsen OH, et al. pcaGoPromoter—an R package for biological and regulatory interpretation of principal components in genome-wide gene expression data. *PLoS One*. 2012;7:e32394.
24. Kress E, Skah S, Sirakov M, et al. Cooperation between the thyroid hormone receptor TR α 1 and the WNT pathway in the induction of intestinal tumorigenesis. *Gastroenterology*. 2010;138:1863–1874.
25. Frasca F, Pandini G, Scalia P, et al. Insulin receptor isoform A, a newly recognized, high-affinity insulin-like growth factor II receptor in fetal and cancer cells. *Mol Cell Biol*. 1999;19:3278–3288.
26. Buxadé M, Parra JL, Rousseau S, et al. The Mnk1s are novel components in the control of TNF α biosynthesis and phosphorylate and regulate hnRNP A1. *Immunity*. 2005;23:177–189.
27. Kjellerup RB, Kragballe K, Iversen L, et al. Pro-inflammatory cytokine release in keratinocytes is mediated through the MAPK signal-integrating kinases. *Exp Dermatol*. 2008;17:498–504.
28. Fan S, Ramalingam SS, Kauh J, et al. Phosphorylated eukaryotic translation initiation factor 4 (eIF4E) is elevated in human cancer tissues. *Cancer Biol Ther*. 2009;8:1463–1469.
29. Berkel HJ, Turbat-Herrera EA, Shi R, et al. Expression of the translation initiation factor eIF4E in the polyp-cancer sequence in the colon. *Cancer Epidemiol Biomarkers Prev*. 2001;10:663–666.
30. Rosenwald IB, Chen JJ, Wang S, et al. Upregulation of protein synthesis initiation factor eIF-4E is an early event during colon carcinogenesis. *Oncogene*. 1999;18:2507–2517.
31. Wendel HG, Silva RLA, Malina A, et al. Dissecting eIF4E action in tumorigenesis. *Genes Dev*. 2007;21:3232–3237.
32. Ueda T, Watanabe-Fukunaga R, Fukuyama H, et al. Mnk2 and Mnk1 are essential for constitutive and inducible phosphorylation of eukaryotic initiation factor 4E but not for cell growth or development. *Mol Cell Biol*. 2004;24:6539–6549.
33. Kligys KR, Wu Y, Hopkinson SB, et al. α 6 β 4 integrin, a master regulator of expression of integrins in human keratinocytes. *J Biol Chem*. 2012;287:17975–17984.
34. Spenlé C, Hussenet T, Lacroute J, et al. Dysregulation of laminins in intestinal inflammation. *Pathol Biol (Paris)*. 2012;60:41–47.
35. Arpin M, Chirivino D, Naba A, et al. Emerging role for ERM proteins in cell adhesion and migration. *Cell Adh Migr*. 2011;5:199–206.
36. Degen M, Natarajan E, Barron P, et al. MAPK/ERK-dependent translation factor hyperactivation and dysregulated laminin γ 2 expression in oral dysplasia and squamous cell carcinoma. *Am J Pathol*. 2012;180:2462–2478.
37. Mercurio AM, Rabinovitz I, Shaw LM. The α 6 β 4 integrin and epithelial cell migration. *Curr Opin Cell Biol*. 2001;13:541–545.
38. Guess CM, Quaranta V. Defining the role of laminin-332 in carcinoma. *Matrix Biol*. 2009;28:445–455.
39. Guess CM, Lafleur BJ, Weidow BL, et al. A decreased ratio of laminin-332 β 3 to γ 2 subunit mRNA is associated with poor prognosis in colon cancer. *Cancer Epidemiol Biomarkers Prev*. 2009;18:1584–1590.
40. Patara M, Santos EMM, de Coudry RA, et al. Ezrin expression as a prognostic marker in colorectal adenocarcinoma. *Pathol Oncol Res*. 2011;17:827–833.
41. Segditsas S, Sieber O, Deheragoda M, et al. Putative direct and indirect Wnt targets identified through consistent gene expression changes in APC-mutant intestinal adenomas from humans and mice. *Hum Mol Genet*. 2008;17:3864–3875.
42. Chen Y, Gelfond JAL, McManus LM, et al. Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC Genomics*. 2009;10:407.
43. Takahashi H, Lassmann T, Murata M, et al. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc*. 2012;7:542–561.